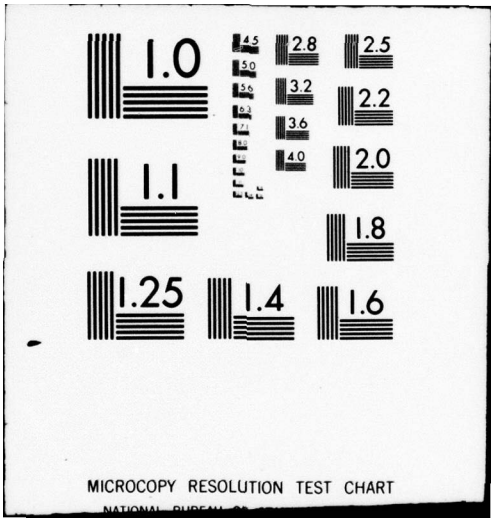
AD-A062 814 MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTE--ETC F/G 12/1
THE INTERPRETATION OF STRUCTURE FROM MOTION.(U)
OCT 76 S ULLMAN
UNCLASSIFIED AI-M-476 N00014-75-C-0643
NL

END
DATE FILMED
2-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
8.0
9.0
MICROCOPY RESOLUTION TEST CHART

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

12

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| | |
|---|---|
| 1. REPORT NUMBER: AIM 476 | 2. GOVT ACCESSION NO. |
| 3. RECIPIENT'S CATALOG NUMBER | |
| 4. TITLE (and Subtitle): The Interpretation of Structure from Motion. | 5. TYPE OF REPORT & PERIOD COVERED: memorandum rept. |
| | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s): Shimon Ullman | 8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0643 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, Massachusetts 02139 | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, 1400 Wilson Blvd, Arlington, Virginia 22209 | 12. REPORT DATE: October 1976 |
| | 13. NUMBER OF PAGES: 35 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of Naval Research, Information Systems, Arlington, Virginia 22217 | 15. SECURITY CLASS. (of this report): UNCLASSIFIED |
| | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

AI-M-476

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None



19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

motion perception
Three-dimensions
Interpretation
scene analysis
object reconstruction
visual motion

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The interpretation of structure from motion is examined from a computational point of view. The question addressed is how the 3-D structure and motion of objects can be inferred from the 2-D transformations of their projected images when no 3-D information is conveyed by the individual projections. The following scheme is proposed: (i) Divide the image into groups of 4 elements each. (ii) Test each group for a rigid interpretation. (iii) Combine the results obtained in (ii). It is shown that this scheme will

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601 | UNCLASSIFIED SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. A correctly decompose scenes containing arbitrary rigid objects in motion, recovering their 3-D structure and motion. The analysis is based primarily on the "structure from motion" theorem which which states that the structure of 4 non-coplanar points is recoverable from 3 orthographic projections. The interpretation scheme is extended to cover perspective projections, and its psychological relevance is discussed.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

A.I. MEMO 476 October 1976

# THE INTERPRETATION OF STRUCTURE FROM MOTION

S. Ullman

Abstract: The interpretation of structure from motion is examined from a computational point of view. The question addressed is how the 3-D structure and motion of objects can be inferred from the 2-D transformations of their projected images when no 3-D information is conveyed by the individual projections.
The following scheme is proposed: (i) Divide the image into groups of 4 elements each. (ii) Test each group for a rigid interpretation. (iii) Combine the results obtained in (ii).
It is shown that this scheme will correctly decompose scenes containing arbitrary rigid objects in motion, recovering their 3-D structure and motion. The analysis is based primarily on the "structure from motion" theorem which states that the structure of 4 non-coplanar points is recoverable from 3 orthographic projections. The interpretation scheme is extended to cover perspective projections, and its psychological relevance is discussed.

This paper is also to appear in the *Proceedings of the Royal Society of London.*

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Project Agency of the Department of Defence under Office of Naval Research contract N00014-75-C-0643.

# THE INTERPRETATION OF STRUCTURE FROM MOTION

## *0. Introduction*

In the course of visual motion perception the changing two-dimensional image is interpreted in terms of objects, their three-dimensional shape, and their motion through space. The remarkable fact that this interpretation requires neither familiarity with, nor recognition of, the viewed objects, was demonstrated by Wallach and O'Connell [1953] in the study of what they have termed the "Kinetic Depth Effect." These experiments showed that the three-dimensional (3-D) structure of unfamiliar objects in motion can be perceived from their orthographic shadow projection. This holds true even when each static view of the object is unrecognizable, and produces no 3-D impression. The original kinetic depth experiments employed primarily wireframe objects whose projection consisted of a connected set of line segments. Later studies [e.g. Wallach and O'Connell, 1953; White and Mueser, 1960; Green, 1961; Braunstein, 1962; Johansson, 1974, 1975] have established that 3-D structure can be perceived from displays consisting of unconnected elements in motion. (The term "elements" will be used to denote any identifiable feature points, such as isolated points, terminations of line segments, or texture elements.) Such displays typically used a small number of elements (one in von Hofsten [1974], two in Borjesson & von Hofsten [1972], three in Borjesson & von Hofsten, [1973], up to six in Braunstein [1962]), or with elements confined to planar surfaces [Gibson & Gibson, 1957; von Fieandt & Gibson, 1959; Gibson *et al* 1959; Gibson, 1965]. In the next section, a demonstration that extends the above demonstrations somewhat by using a large number of points arranged in two non-planar configurations is described. It will exemplify the perception of structure from motion, and will help in formulating the computational problems underlying this perception.

*1. The two cylinders demonstration*

The orthographic projection of two coaxial cylinders was presented on a computer-controlled CRT screen. Each cylinder was defined solely by 100 points lying on its surface. The common axis of the two cylinders was vertical, as diagrammed in Figure 1. The 3-D coordinates of the points were stored in the computer's memory; their orthographic projection on the frontal plane was computed and presented on the screen. The imaginary cylinders were then rotated (up to about 10 degrees at a time), their new projection was computed and displayed on the screen (presentation time being 100 msec. with 40 msec. ISI). In the projected image the dots increased in density at the edges of each cylinder, but in the image of the two cylinders the variations in density where complex and ineffective in revealing the 3-D structure of the displayed objects. Each single static view thus appeared to be an almost random collection of points. However, when the changing projection was viewed, the elements in motion across the screen were perceived as two rotating cylinders whose shapes and angles of rotation were easily determined. Both the segmentation of the scene into objects and the 3-D interpretation were based in this case on motion alone, since each single view contained no information concerning the segmentation or the structure. Each frame in the presentation was an unfamiliar, unrecognizable view of the two cylinders, indicating that familiarity and recognition are not prerequisites for the interpretation of motion.

Two restrictions of the above demonstration are noteworthy. Firstly, the rotation axis employed in the demonstration was fixed in space throughout the motion. However, similar demonstrations in which the orientation of the rotation axis changed abruply (by 30 deg. and more) after each frame were examined as well, and the 3-D structure was still perceptible (c.f. the "tumbling motion" in [Green, 1961]). Secondly, the demonstration employed discrete stimuli in apparent motion. However, this appears to be immaterial to the interpretation process. Three-dimensional structure can be perceived from both continuous and apparent motion, and the subsequent analysis will be applicable to both.

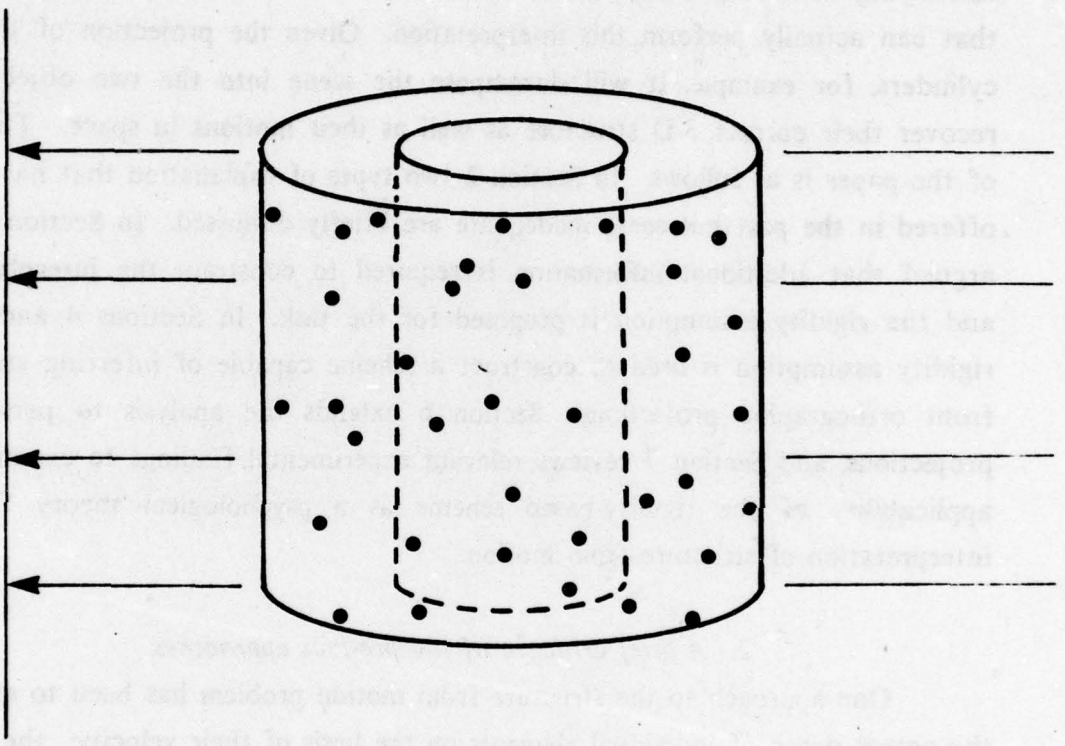


**Figure 1. The projection of the two cylinders: a side view.**
**(The outlines of the cylinders were not presented in the actual display.)**

The objective of this paper is to examine the computational problems underlying the interpretation of structure from motion, and to develop a scheme that can actually perform this interpretation. Given the projection of the two cylinders, for example, it will decompose the scene into the two objects and recover their correct 3-D structure as well as their motions in space. The plan of the paper is as follows. In Section 2 two types of explanation that have been offered in the past but seem inadequate are briefly discussed. In Section 3 it is argued that additional information is required to constrain the interpretation, and the rigidity assumption is proposed for the task. In Sections 4 and 5 the rigidity assumption is used to construct a scheme capable of inferring structure from orthographic projections. Section 6 extends the analysis to perspective projections, and Section 7 reviews relevant experimental findings to examine the applicability of the rigidity-based scheme as a psychological theory for the interpretation of structure from motion.

## *2. A brief criticism of two previous approaches.*

One approach to the structure from motion problem has been to estimate the actual depth of individual elements on the basis of their velocity: the higher the velocity of an element, the closer it is [Helmholtz, 1910; Braunstein, 1962; Hershberger & Starzec, 1974]. According to this view, the recovery of structure from motion is analogous to depth perception through stereopsis, with successive frames substituting for adjacent images, and displacement values playing the role of binocular disparity. Outside some special situations (e.g. pure translation of the observer in a stationary environment), this scheme cannot be correct since in the general case displacement values (or, equivalently, velocities), are not indicative of depth (e.g. in the two cylinders example). The scene might include objects moving in different directions and at various speeds with no correlation between velocity (or displacement) and depth. In the two cylinders example, while velocity cannot serve as an indication of depth, within each cylinder velocity changes in accordance with depth. It might therefore be suggested that

the grouping of the elements into bodies should precede the depth analysis. Possibly this consideration was one reason why grouping by motion has been studied as a problem on its own. The Gestaltists, for example, had the notion of "grouping by common fate" which included grouping by common velocity. Potter [1974] used a similar criterion: elements i and j with velocities $v_i$ and $v_j$ respectively are grouped if $|v_i - v_j|$ is less than some pre-determined threshold. The two cylinders illustrate the difficulties involved in grouping by motion. Each cylinder contains points spanning a range of velocities, while points having exactly the same speed belonged to different objects.

A different explanation for the interpretation of structure from motion was offered in the original study of the kinetic depth effect [Wallach & O'Connell, 1953] as well as in later studies [Wallach *et al.*, 1956; Jansson and Johansson 1973]. The ability to perceive structure from motion was accounted for in terms of an "effect" produced by lines and contours that change simultaneously in both length and orientation. If only actual lines in the image were considered, the account is manifestly false, since the structure of unconnected dots can be recovered through their motion. Imaginary lines connecting identifiable points were therefore admitted as well [Wallach & O'Connell, 1953]. But the resulting condition (i.e. that the perception of 3-D structure is produced by lines, virtual lines and contours that change in both length and orientation) is grossly insufficient. Consider for example the random motion of unconnected elements in the frontal plane. The virtual lines between them change constantly in both length and orientation but no coherent 3-D structure is perceived. The above condition is also necessary in a trivial sense only: The only 2-D transformations of the image that violate Wallach and O'Connell's condition are rigid transformations (of the image, not of the 3-D objects) and uniform scaling. But if the structure of a 3-D object is not recoverable from a single projection, a uniform displacement or scaling of the projection are insufficient to reveal the unknown structure.

### *3. The rigidity assumption*

The fundamental problem underlying the interpretation of structure from motion is the ambiguity of the interpretation. That is, there is no unique structure and motion consistent with a given two-dimensional (2-D) transformation [Eriksson 1973]. In the two cylinders demonstration for instance, the elements which move on the 2-D screen are perceived as elements on 3-D cylinders in rotation. Furthermore, these two interpretations (the planar and the two cylinders) are not the only ones consistent with the displayed transformation. They are but two of an infinite number of motions of the elements that will produce the same 2-D projection.

To cope with this indeterminacy of structure, the interpretation scheme must incorporate some internal set of constraints that rule out most of the possible 3-D interpretations and force a unique solution, which in most real cases is also the veridical one. To be restrictive enough on the one hand, and not misleading on the other, they can be incorporated in an interpretation scheme only if they meet the following requirements. Firstly, they should reduce the number of solutions to a unique one, at least in most cases, and secondly, the constraints should be plausible, in the sense that they almost always hold true in the environment.

The constraint I propose for the interpretation of structure from motion is what I shall call the *rigidity assumption*:

***Any set of elements undergoing a two dimensional transformation which has a unique interpretation as a rigid body moving in space should be interpreted as such a body in motion.***

In giving priority to rigid interpretations I follow several researchers [e.g. Wallach & O'Connell, 1953; Gibson & Gibson, 1957; Green, 1961; Hay, 1966; Johansson, 1975; and a related "three-dimensionality principle" by Johansson, 1964, 1975] who observed that rigidity seems to play a special role in motion perception. However, this "bias for rigidity" is only the starting point of the analysis. The rigidity assumption means that this bias does more than to reflect

a concern with rigid objects. The assumption capitalizes on properties of the physical world to arrive at the correct interpretation of under-determined stimuli. The next step in the analysis is therefore to show how the rigidity assumption can be incorporated in an interpretation scheme that will correctly infer structure from motion.

### *4. Incorporating the rigidity assumption in an interpretation scheme*

To use the rigidity assumption, the interpretation scheme must be able to check whether a given collection of moving elements has a unique rigid interpretation. The interpretation can then proceed by submitting sub-collections of the elements in the scene to the following *rigidity test*:

***Does this collection have a unique interpretation as a rigid body moving in space?***

If the answer is negative (either because there is more than a single rigid interpretation or because there are none), then no specific structure is attached to the elements. If the answer is positive, the unique structure discovered is imposed upon the elements.

The rigidity assumption as stated in the previous section requires that the test be administered to small sub-collections of the elements in the scene. This would be necessary if, for example, the scene were composed of several objects participating in different motions. If we test all the elements in the scene at once for rigidity, the test might fail simply because the elements belong to more than a single object. It follows that the rigidity test must be administered to what I shall term a *nucleus* of elements, namely, a minimal number of elements which is still sufficient to determine their structure uniquely. We shall shortly see what this nucleus is.

The above rigidity test raises two immediate problems. The first one is whether the test is computable. Namely, is there a procedure for deciding whether a collection of moving elements has a unique rigid interpretation, and for actually determining that structure. The second problem is whether the

proposed procedure will result in the correct interpretation of the input projections. We shall address the second of these problems first, by examining the possible ways in which the interpretation procedure might go wrong. One possibility of error arises when the rigidity test answers "yes" when it should have answered "no", the second when it answers "no" instead of "yes." The first error involves "false targets": points that actually move independently of each other that just happen to have a unique interpretation as a rigid body in motion. In this case the interpretation scheme will produce the false structure it has stumbled upon. The second kind of error results from "phantom structures": points that are the actual projection of a certain moving object that are also compatible with the projection of a different object in motion. Because of the additional "phantom structure" the solution would fail to be unique, and consequently no structure would be assigned to the moving elements.

The power of the rigidity-based interpretation scheme stems from the fact that the probabilities of committing a misinterpretation of either type are negligible. We shall examine first the phantom structures problems and show that under simple assumptions they are impossible. That is, given the 2-D orthographic projection of a certain object in motion, there is no other object, or a different rigid motion, compatible with the given projection. This claim follows from a theorem concerning rigid objects, which I shall call the "structure from motion" theorem.

***The structure from motion theorem:***

***Given three distinct orthographic views of four non-coplanar points in a rigid configuration, the structure and motion compatible with the three views are uniquely determined.***

This theorem was originally stated and proved by Ullman [1977, Appendix 1] for five points. D. Fremlin [1977, personal communication] showed that the requirement can be relaxed to four points. A proof combining [Ullman 1977] and [Fremlin 1977] is given in the appendix.

The views in the structure-from-motion theorem are obtained by

orthographic projection. As demonstrated e.g. by the kinetic depth experiments and by the two cylinders demonstration, the human visual system can infer structure from orthographic projections, and this is the case we shall examine first. In Section 6 the results will be extended to cover perspective projections as well.

The theorem has two implications for the interpretation of structure from motion. Firstly, it establishes that 3-D structure can be recovered from as few as four points in three views. This is, then, the minimal nucleus on which the interpretation scheme can operate. Secondly, the fact that the structure is uniquely determined implies that phantom structures are impossible. Hence, this type of misinterpretation is ruled out. The second type of misinterpretation I have mentioned was false targets. It can be shown, however, that in our 3-D world false targets are highly unlikely: the probability that three views of four points not moving rigidly together will admit a rigid interpretation is low. In fact, the probability is zero, provided that the position of the points is given with complete accuracy. (The argument supporting this claim is statistical, and is given by Ullman [1977, Appendix 1].)

Of the two possible misinterpretations listed above, phantom structures are impossible while false targets have zero probability of occurrence. Consequently, the interpretation scheme is virtually immune from misinterpretation. It should be noted that since orthographic projections are employed, the object is determined only up to reflection about the frontal plane. This ambiguity is inherent, since an object rotating by some angle — and its mirror image rotating by -— have the same orthographic projections. Similarly, the absolute distance to the object and its translation in depth cannot be recovered from its orthographic projection. The interpretation in the orthographic case thus gives one: a) the decomposition of the scene into objects, b) the 3-D structure of these objects up to reflection, and c) the motions of the objects (the relation between the initial and final position and orientation) up to translation in depth.

The formulation of the structure-from-motion theorem in terms of three

distinct views does not imply that the motion of the input image has to be discrete (as opposed to continuous). If a continuous motion extends long enough to contain three distinct views (and what qualifies as distinct depends on the accuracy of the interpreting system), then it contains sufficient information for a unique interpretation. The theorem states this mathematical fact without implying how this information should be extracted.

*Summary of the main principles:* The main principles underlying the structure-from-motion interpretation scheme can be summarized by dividing the interpretation problem into two main components. The first sub-problem is that the 3-D structure and motion are under-determined by the projected 2-D transformations. This difficulty was overcome by incorporating the rigidity assumption as an internal constraint. The second problem in recovering the original motion is that the 2-D transformations in a given scene might be induced by several objects, engaged in different 3-D motions. This difficulty was avoided by restricting the interpretation of motion to nuclei of elements which would generally belong to a single object.

### *5. Implementation of the scheme and its application to large collections of elements*

The proof of the structure-from-motion theorem is constructive, offering a way of devising a scheme that actually recovers structure from motion. Such a scheme has been implemented at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Some comments regarding the implementation are found in the appendix, but a few conclusions are worth mentioning here.

*Planar objects:* The structure-from-motion theorem guarantees a unique solution for three views of four non-coplanar elements. This does not mean, of course, that the non-coplanarity has to be known before the rigidity test is applied.

Note also that the non-coplanarity condition is sufficient, not necessary: four coplanar elements might have many solutions or just a single one, depending on the initial orientation of the planar object and its subsequent rotations in space. If they have a unique solution, the structure will be recovered by the algorithm. Otherwise, the fact that they lie on a single plane will be established, but its orientation and rotation will remain ambiguous. A similar situation arises when the viewed object is composed of only three points. Some three-points configurations are uniquely determined by three views, others are not. The rigidity based algorithm can be applied to three views of three points, in which case it is no longer guaranteed to yield a unique solution. However, if the interpretation happens to be unique, it will be discovered by the algorithm.

*Number of points vs. accuracy trade-off:* In the algorithm, there is a possible trade-off between the number of points (or views) used and the accuracy required from the input and the computation. If the input data is given with high accuracy and if the computation process is similarly accurate, then four elements are sufficient. A less accurate device (like, perhaps, our visual system) might require more elements (or more views) for a reliable and accurate interpretation.

*Application to large collections of points:* Since real scenes typically contain a large number of elements, the complexity of the computations involved in the interpretation process needs to be examined. The question is whether the computation remains manageable as the number of elements grows into the hundreds or the thousands. The answer is that in realistic scenes the amount of computation required is expected to grow only about linearly with the number of points. Furthermore, the process can be carried out mostly in parallel so that the computation time can be largely independent of the number of points.

To examine the many-elements situation, assume that there are $N$ elements in the image and $K$ objects. We can divide the set of $N$ elements into $N/4$ groups, each containing four neighboring elements and carry out the interpretation scheme on each of the groups. The computations on the different

groups are independent of each other and could be performed in parallel. For all groups interior to the image of an object, namely those in which all four elements belong to the same object, the rigidity test will succeed and the structure will be discovered. The argument now depends on the expectation that the points which comprise a given object will not be distributed randomly over the entire scene. In the case of real, opaque objects, it is expected that each object will have at least one interior group. (The case involving a few transparent objects is somewhat more complex but not unmanageably so, see below.) The first step will thus yield for each of the $K$ objects a set of interior points whose structure and motion are determined, and a set of boundary points which are as yet undetermined. The next step checks, for each of the remaining boundary points, which of the $K$ objects it belongs to. This step can also be executed for all the points in parallel.

*The two cylinders and non-opaque objects:* The case of non-opaque objects complicates the computation since points chosen locally can belong to different objects, one behind the other. However, if the number of visible objects at each location is small, the increase in complexity is limited. For example, the structure-from-motion algorithm has been applied to the two cylinders display, where two objects are visible, one inside the other. In the central region (where the two cylinders overlap) most of the groups of four points selected at random contain elements of both cylinders and therefore do not have a rigid interpretation. However, 1/8 of the groups do have a rigid interpretation (1/16 for each cylinder). Also, all the groups in the non-overlapping region belong to the bigger cylinder and have a unique interpretation. Thus, after the first step, about 1/8 of the points are assigned a 3-D structure. A second step completes the interpretation as explained above for boundary points.

## *6. Perspective projections*

In inferring 3-D structure from perspective rather than orthographic projections, one possible approach is to modify the foregoing analysis to reflect the different underlying geometry. That is to say, the interpretation scheme will test 2-D transformations for their compatibility with the perspective projection of a rigid 3-D configuration in motion. This approach was examined by Ullman [1977, Section 3.2], and was found to suffer from three shortcomings.

1. Perspective effects are often small, hence a procedure that relies on them depends on highly accurate input and is sensitive to small errors. This problem is especially acute if a rigidity-based interpretation is to be performed locally, since perspective effects diminish with the size of objects.
2. The computations increase considerably in complexity.
3. The performance of the resulting method differs from human performance in a way that makes it unattractive as a basis for a psychological theory.

In this section we shall therefore follow a different approach, based on the orthographic structure-from-motion scheme. It will be shown that this scheme can be applied to yield approximate results in the perspective case, leading to a scheme that is comparable with human performance in both its capacity and its limitations.

### *6.1 The polar-parallel structure-from-motion interpretation scheme*

If perspective projections are used instead of orthographic ones, and if the object is sufficiently "far away", its perspective projections can be viewed as slightly distorted orthographic projections. In such a situation, the structure computation used in the orthographic case can be used to provide approximate results in the perspective case as well. The term "far away" means that the difference between the distances from the observer to the points in question is small compared to the distances themselves. That is, if $y_i$ and $y_j$ are the distances of two points from the observer, the points are "far away" if the value of $|y_i - y_j|$ is much smaller than $y_i$ and $y_j$. Such a condition can hold regardless

of the actual distance of the object from the observer, if two requirements are met; firstly, the object has densely distributed visible points, and secondly, it is continuous so that for nearby points in the image, the separation of their spatial sources is also small. In such a case we can take advantage of the locality of the structure computation, i.e. that only four points are needed. If the object in question obeys the above two requirements, then by limiting the field of view to four nearby points at a time, the interpretation is performed on a "far away" object and therefore the local structure can be recovered. Applications of the orthographic structure-from-motion algorithm to perspective projections showed that when the relative depth of the points is less than 10% of the absolute depth, the error in the computed structure and motion is usually also limited to less then 10%.

Applying the orthographic structure-from-motion scheme locally and then combining the results is different from analyzing the entire image at once as a single orthographic projection. In orthographic projection there is a single axis of projection common to all the points. In applying the orthographic scheme to small neighborhoods, while each neighborhood is treated as an orthographic projection, the axis of projection changes from one neighborhood to the other. To interpret the structure of an object using the orthographic method, we first divide it up into many regions containing about four elements each. We then analyze each region as if it were obtained from a orthographic projection whose axis is along the line from the eye to the center of the region in question. This way of analyzing perspective projections will be called the *polar-parallel* method.

### *6.2 Determining the structure uniquely.*

In orthographic projection the interpretation is determined up to a reflection about the frontal plane. The polar-parallel scheme does not share this ambiguity, for although the structure is determined locally only up to a reflection, global consistency requirements make it possible to distinguish between the true and the locally reflected structures. Figure 2 helps to illustrate

this point. In 2a an object $O$ is subject to a orthographic projection. The two possible interpretations are depicted: the correct one $I$ (solid line) and the reflected one $I'$ (dotted line). In Figure 2b the same object is subject to a perspective projection. The recovery of the structure is performed locally by the polar-parallel scheme on regions like $O_1$ and $O_2$ in 2b. The figure presents the interpretations of two local regions. Each local operation determines the structure up to reflection: $I_1$ and $I'_1$ for $O_1$, and $I_2$, $I'_2$ for $O_2$. Since the direction of the projection changes from $O_1$ to $O_2$, the directions of the reflected structures changes as well. While $I_1$ and $I_2$, the correct interpretations of $O_1$ and $O_2$, are parallel to each other, this relation no longer holds for $I'_1$ and $I'_2$.

While Figure 2 demonstrates the distinction between orthographic and polar-parallel interpretations, it does not reveal how the incorrect structure is ruled out. For all the observer knows, either interpretion in 2b might be the true one. The correct interpretation can, however, be singled out by the requirement for consistency of the axes of rotation across the object. Like the computed structure, under orthographic projection the axis of rotation between two succesive views has two possible solutions: the true one, and a reflected one. While the true orientation remains constant over the entire object, the orientation of the reflected axis changes from one region of the object to another. The requirement that a rigid body in rotation has a unique axis of rotation imposes a unique interpretation on the structure as well. Once the axis of rotation is uniquely determined, the structure of the object is also recovered uniquely. The discovery of the true axis is facilitated when the object increases in size. The reason is that when the axis of rotation is recovered for two distinct regions of the object, they will have one axis (the true one) in common, while the reflected ones will be separated by an amount that increases with the angular separation between the two regions. It follows that the maximum separation between two false axes is achieved for two regions of the object in question which are as separated as possible. Consequently, the larger the angular extent of the object, the easier and more reliable is the discrimination between

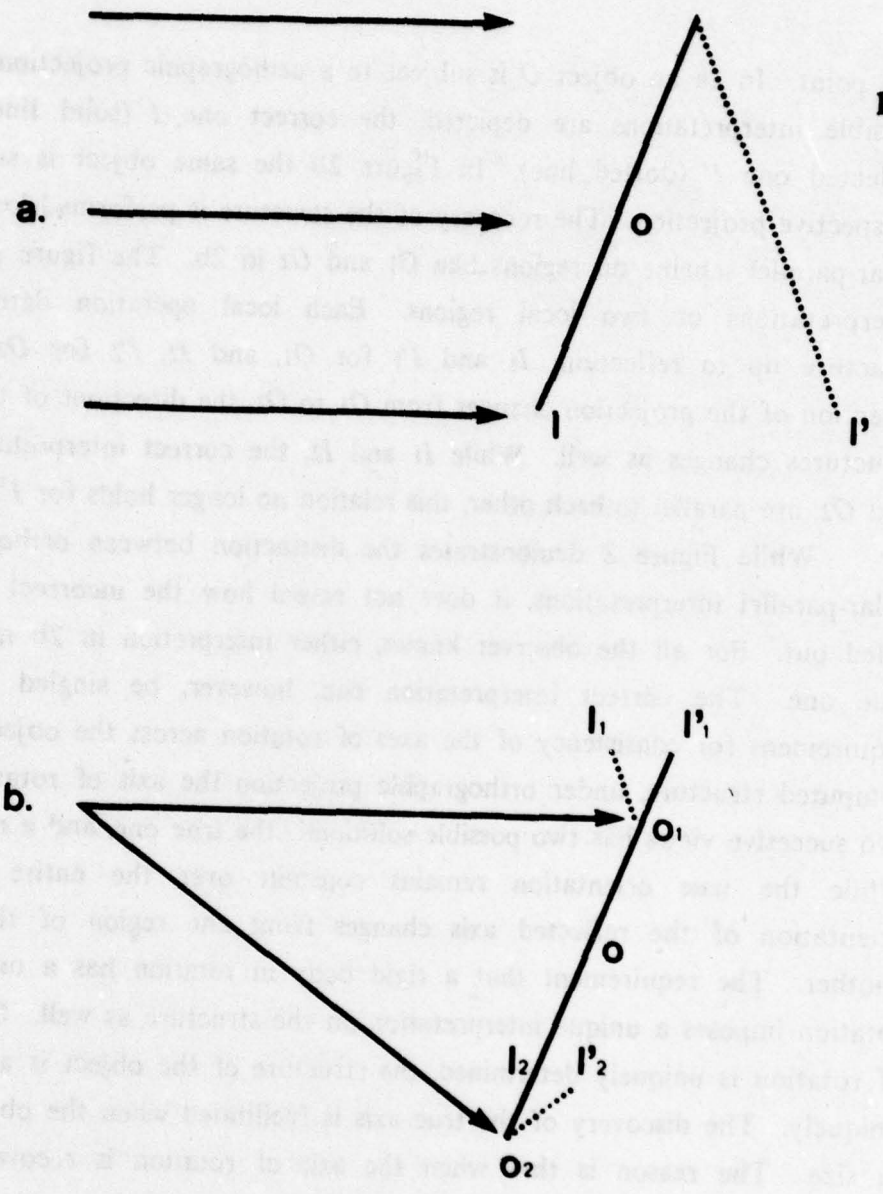



**Figure 2. The ambiguity of interpretation in the case of orthographic (2a) and polar-parallel (2b) projections. In 2a the ambiguity is global. In 2b the ambiguity is local and can be resolved.**

the true and the spurious axes. Another factor that facilitates the choice of a unique structure is the object's texture. The smaller the separation between identifiable elements on the object, the more accurate is the recovery of the true axis, and the easier is its separation from the spurious ones.

### *7. Psychological relevance*

In this section I shall review experimental findings that suggest that principles similar to those underlying the polar-parallel structure-from-motion interpretation scheme are used by our visual system to infer structure from motion.

#### *7.1 orthographic projections*

1. *The need for identifiable elements:* The structure-from-motion interpretation scheme helps to explain a phenomenon which Wallach and O'Connell in their original study of the "kinetic depth effect" considered a baffling enigma. When the objects used in their experiments were smoothly curved, so that their shadows did not correspond to identifiable, traceable, 3-D points, the 3-D structure was impossible to recover from the projection:

"Curved contours which are deformed without displaying a form feature which identifies a specific point along the curve, are seen as distorting [rather than moving in depth] ... This peculiarity is in disagreement with our description of the kinetic depth effect and has delayed our work for years." [Wallach & O'Connell 1953, p. 209].

Failure to recover structure from motion under these circumstances is to be expected from the structure-from-motion interpretation scheme. According to this scheme the different views are not merely associated somehow into a single whole. Rather, the motion of the individual elements is checked for consistency with a motion of a rigid body. Consequently, the interpretation scheme will fail in the absence of identifiable elements which can be reliably traced throughout the 2-D transformation.

*Number of points*: A major feature of the structure-from-motion interpretation scheme is that the structure can be recovered from a small number of elements. Four non-coplanar points are always sufficient, and three sometimes suffice too, especially if more than three views are provided. For the human visual system the interpretation does not seem to be an all-or-none phenomenon. The accuracy and the stability of the perceived structure increases with the number of elements and views [Green, 1961; Braunstein, 1962]. However the minimum needed for correct interpretation seems to be comparable with the structure-from-motion interpretation scheme: the correct structure of as few as three elements in motion is sometimes perceivable.

*Reversals:* As explained in the structure-from-motion theorem, the interpretation in the pure orthographic case is determined up to reflection about the frontal plane. The viewed object may thus undergo a depth reversal which must be accompanied by a switch in the observed sense of rotation. Many experiments [e.g. Wallach & O'Connell, 1953; Wallach, O'Connell & Neisser, 1953; White & Mueser 1960], have established that objects viewed in orthographic projection do undergo spontaneous depth reversals accompanied by switches in the observed sense of direction.

*Two points in motion:* Some cases in which the structure fails to be unique can still be interpreted if additional assumptions are made. The motion of a single line segment (or of its two endpoints) provides an example of such a case. The two points can always be interpreted as the endpoints of a rigid rod whose orientation and rotation are not uniquely determined. However, the two missing variables are related: once the orientation is known the rotation is determined and vice versa. Johansson and Jansson [1968] examined the perception of a single line in motion. Their results concerning the judged orientation of the rod show a tendency to assume that the rod lies in the frontal plane at the moment when it has its maximal extension.

Note that this "maximal extension" assumption cannot be a part of the interpretation scheme in general. Consider the two cylinders example and assume that the total rotation observed is less than 90 degreees. In this case the orientations of pairs of points are inconsistent with the "maximum extension" assumption, and the correct structure, not the one implied by the assumption, is perceived. (Such an experiment was performed by Wallach & O'Connell [1953]. Rotation through 42 degrees was sufficient to reveal the structure of the hidden objects.) The maximal extension assumption can only serve as a rough and unreliable "last resort" when the general interpretation scheme (which requires uniqueness) fails. Though unreliable in general, this assumption is still the safest one in the impoverished situation of only two elements. It seems that the human visual system tends to use it under such conditions, but without placing much confidence on the results: our 3-D perception of the rotating rod is usually weak and unstable.

*Planar objects:* Flat objects do not obey the non-planarity requirement, and so the correct recovery of their structure is not guaranteed. Although for most planar objects the structure will nevertheless be recoverable, some cases are inherently ambiguous. For example, let *l* be the intersection of the object plane with the the image plane, and assume that the object rotates about an axis parallel to *l*. In this configuration the fact that the points are coplanar can be established, but the initial orientation of the plane and its subsequent rotations remain as dependent but unknown variables. Gibson & Gibson [1957] found that planar objects in orthographic projection are indeed ambiguous under the described rotation. (By analogy with the two points case, it seems that in this under-determined situation humans exhibit some tendency to interpret either the initial or the maximal extension position of the plane as frontal.) In contrast with the above condition, the structure of a tilted plane (one that does not pass through the vertical axis) in rotation about the vertical axis is recoverable by the structure-from-motion scheme as well as by human observers.

*Absolute and relative depth*: The structure-from-motion interpretation scheme recovers the structure of rigid objects. In contrast with the structure which involves relative depth, the absolute depth is not recovered. Furthermore, the relative depth of two objects which move independently of each other cannot be determined. In experiments carried out by Gibson *et al* [1959], subjects were able to determine the correct slant of a projected plane in motion, while the absolute distance estimates varied from 3 inches to 5 miles. When several planar objects were presented, their separation in depth was perceivable when they moved rigidly together [Gibson, 1957] but not when their motions were independent [Gibson *et al*, 1959].

### *7.2 Perspective projections*

*Uniqueness of the solution: effects of size, texture, and tilt.* It has been mentioned that favorable conditions for the polar-parallel scheme to distinguish between the correct interpretation and its mirror image include large angular extension and dense texture. For planar objects in rotation the unique interpretation is also facilitated when the plane is tilted as already dicussed. These expectations are corroborated by findings concerning the Ames trapezoid window [Ames, 1951]. The probability of perceiving the correct orientation and rotation of the window depend on its size, texture, and tilt [ Zegers, 1964; Epstein *et al*, 1968].

*Unique direction of rotation and the "motion parallax" cue.* When the perspective projection of a rotating object is viewed, the correct direction of rotation is usually perceived, in contrast with the spontaneous reversals that characterize orthographic projections. Several attempts have been made [Braunstein, 1962; Hershberger & Sarzec, 1974] at using velocity differences to resolve the structure and rotation ambiguity by suggesting that nearer points can be distinguished from farther ones by their greater velocity. The use of angular velocity differences as an indication of relative depth, usually referred to as "motion

parallax", relies on the following relation between angular velocity and distance. When an element moves in space with velocity $v$ at angle of — deg. with the observer's line of sight and at a distance $r$ from his eye, the angular velocity $w$ of the element relative to the observer is given by: $w = v \sin— / r$. A procedure can be said to use the motion parallax cue if it uses $w$ and $v \sin—$ to determine $r$. A straightforward example is the case of a uniform displacement of the environment, e.g. the one caused by the observer's translation. Along a given line of sight $v \sin—$ is constant, and therefore the angular velocity of elements along a line are inversely proportional to their distance from the observer. Motion parallax "cues" were advanced as explanations for humans' ability to distinguish the true sense of rotation from the confusable one. It seems, however, that the applicability of such cues is too limited to account for this ability, whereas the polar-parallel scheme seems suitable for the task. Consider for instance a rigid rod in rotation about its midpoint. Although the velocities of its two endpoints are equal in magnitude (and opposite in sign) the inverse relation between speed and distance no longer holds. Since the angles — between the velocity vectors of the elements and their respective lines of sights are different, the difference between their angular velocities becomes a rather complex function of the rod's position [Hershberger, 1967]. It turns out [Ullman 1977, p. 161-2] that the statement "the faster endpoint is the nearer one" is erroneous over half of the cycle time (For example, if the distance to the rod is twice the rod's length, it will be erroneous for 50 out of 90 deg. of rotation). It follows that the parallax cue cannot be reliably used independently of an estimation of the rod's orientation. There are additional severe problems with using the motion parallax cue. It cannot be used when the rod does not rotate about its midpoint, or when the moving elements are not at 180 degrees to each other, or when the rod's motion is not confined to a pure rotation. Other "cues" proposed for determining the rotation direction (e.g. the use of the stagnation points suggested by Hershberger and Urban, [1970] ) are susceptible to the same problems, in particular to the one caused by compound motion i.e. rotation

accompanied by translation The failure of traditional motion parallax cues under compound motion reflects an important difference between them and the structure-from-motion scheme. In the structure-from-motion scheme, segmentation, structure, rotation, and translation, are not treated independently. (unlike [Braunstein, 1962; Gibson 1957; Hay, 1966; Borjesson and von Hofsten, 1972; 1973; Johansson, 1974; Eriksson, 1974] ). If sufficient information (enough points and views) is supplied, the segmentation, structure, rotation, and translation are uniquely determined although none of them is determined by any "cue" in isolation.

*Non-rigid motion*: The structure-from-motion scheme cannot be applied to non-rigid deformations. However, since the interpretation process is local, it is applicable to quasi-rigid motions which approximate locally rigid motion (e.g. "bending" motion, [Jansson & Johansson, 1973] ).

*Few, widely separated points*: For the polar-parallel scheme, the recovery of the 3-D structure from the perspective projections of few (about 4-5) points decreases in accuracy as the perspectivity increases. The perspectivity depends on the ratio between the object's size and its absolute distance from the viewing point. The smaller the perspectivity, the closer is the projection to the orthographic case, and therefore the higher the accuracy of the polar-parallel scheme. In contrast, for a scheme that uses perspective projections directly, large perspective effects should not hinder the interpretation. Braunstein's [1962] findings suggest that the peception of rigidity depends on perspectivity in the expected way. It is strongest for orthographic projection, whereas perspective effects cause perceived distortions in the moving object.

*Structure from receding motion:* In an interpretation scheme that relies on perspective projections directly, the distortions of an image caused by translation in depth provide a rich source of information that makes this type of motion

particularly easy to analyze (c.f. the perspective scheme in [Ullman 1977] ). In the polar-parallel scheme on the other hand, these distortions are viewed as "noise" which impedes the computation and thereby makes receding motion harder to analyze. The recovery of structure from receding motion by the polar-parallel scheme is possible when the object is large and textured, and when its translation in depth is sufficiently large. It seems that the same requirements have to be met for human observers to recover structure from receding motion. Gibson *et al* [1959] have shown that the slant of a receding plane can be judged under conditions highly favorable for the polar-parallel scheme. The planar object tested was highly textured and extended over 82 deg. of visual angle. I have also found that the perception of structure from receding motion becomes impossible under less favourable conditions. The two cylinders of Section 2 were presented in receding motion rather than rotation. The presentation comprised 8 frames. When viewed at a distance of 80 cm. from the screen, it simulated a gradual motion in depth of two cylinders of diameters 25 and 50 millimeters from an initial distance of their common axis of 17 cm. to 62.5 cm. The sequence was run forward (receding motion), backwards (approaching motion), and alternating between both motions in a cycle. Although the contraction and expansion of the image elicited some impression of motion in depth, the structure of the cylinders could not be recovered.

I thank D. Marr for helpful suggestions and comments, and B. Schatz and K. Stevens for careful reading of the manuscript. This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Project Agency of the Department of Defence under Office of Naval Research contract N00014-75-C-0643. The work was also supported by NSF grant 77-07569-MCS.

## REFERENCES

Ames, A. 1951 Visual perception and the rotating trapezoid. *Psycholog. Monographs*, 65(7), Whole No. 324, 1-32.

Borjesson, E. and von Hofsten, C. 1972 Spatial determinents of depth perception in two-dot motion patterns. *Percept. & Psychophys.* 11(4) 263-268.

Borjesson, E. and von Hofsten, C. 1973 Visual perception of motion in depth: Application of a vector model to three dot motion patterns. *Percept. & Psychophys.* 13(2) 169-179.

Braunstein, M. L. 1962 Depth perception in rotation dot patterns: effects of numerosity and perspective. *J. Exp. Psychol.* 64(4), 415-420.

Epstein, W. Jansson, G. and Johansson, G. 1968 Perceived angle of oscillatory motion *Percept. & Psychophys.* 3(1A) 12-16.

Eriksson, E. S. 1973 Distance perception and the ambiguity of visual stimulation: A theoretical note. *Percept. & Psychophys.* 13(3), 379-381.

Fieandt, von Kai and Gibson, J. J. 1959 The senstivity of the eye to two kinds of continuous transformations of a shadow-pattern. *J. Exp. Psychol.* 57, 344-347.

Fremlin, D. 1977 Personal communication.

Gibson, E. J., Gibson, J. J., Smith, O. W. and Flock, H. 1959 Motion parallax as a determinant of perceived depth. *J. Exp. Psychol.* 8(1), 40-51.

Gibson, J. J. 1957 Optical motions and transformations as stimuli for visual perception. *Psychol. review* 64(5), 288-295.

Gibson, J. J. and Gibson, E. J. 1957 Continuous perspective transformations and the perception of rigid motion. *J. Exp. Psychol.* 54(2), 129-138.

Gibson, J. J. 1965 Research on the visual perception of motion and change. In: Spigel, I. [1965] 125-146.

Gibson, J. J. 1968 What gives rise to the perception of motion? *Psychol. Review* 75(4), 335-346.

Green, B. F. 1961 Figure coherence in the kinetic depth effect. *J. Exp. Psychol.* 62(3), 272-282.

Hay, C. J. 1966 Optical motions and space perception - an extention of Gibson's analysis. *Psychol. Review* 73, 550-565.

Helmholtz, H. L. F. von. 1910 *Treatise of physiological optics.* Translated by J. P. Southall, 1925, N.Y. Dover publications.

Hershberger, W. A. 1967 Comment on apparent reversal (oscillation) of rotary motion in depth. *Psychol. Review* 74, 235-238.

Hershberger, W. A. and Urban, D. 1970 Depth perception from motion parallax in one dimensional polar projection: projection versus viewing distance. *J. Exp. Psychol.* 86, 380-383.

Hershberger, W. A. and Starzec, J. J. 1974 Motion-parallax cues in one dimensional polar and orthographic projections. *J. Exp. Psychol.* 103(4), 717-

723.

von Hofsten, C. 1974 Proximal velocity change as a determinant of space perception. *Percept. & Psychophys.* 15(3), 488-494.

Jansson, G. and Johansson, G. 1973 Visual perception of bending motion. *Perception* 2, 321-326

Johansson, G. 1964 Perception of motion and changing form. *Scan. J. Psychol.* 5, 181-208.

Johansson, G. and Jansson, G. 1968 Perceived rotary motion from changes in a straight line. *Percept. & Psychophys.* 4(3), 165-170.

Johansson. G. 1973 Visual perception of biological motion and a model for its analysis *Percept. & Psychophys.* 14(2), 201-211.

Johansson, G. 1974 Visual perception of rotary motion as transformation of conic sections -- a contribution to the theory of visual space perception. *Psychologia* 17, 226-237.

Johansson, G. 1975 Visual motion perception. *Sci. Am.* 232(6), 76-88

Potter, J. 1974 The extraction and utilization of motion in scene description. *Ph.D. Thesis*, University of Wisconsin.

Slocum, R. V. and Hershberger, W. A. 1976 Perceived orientation in depth from line-of-sight movement. *Percept. & Psychophys.* 19(2) 176-182.

Spigel, I. (ed.) 1965 *Readings in the Study of Visually Perceived Motion* New

York: Harper and Row.

Ullman, S. 1977 The Interpretation of Visual Motion. *Ph.D. Thesis*, M.I.T., Department of Electrical Engineering and Computer Science.

Ullman, S. 1978 *The Interpretation of Visual Motion.* Cambridge, M.I.T. Press

Wallach, H. and O'Connell, D. N. 1953 The Kinetic depth effect. *J. Exp. Psychol.* 45(4), 205-217.

Wallach, H., O'Connell, D. N., and Neisser, U. 1953 The memory effect of visual perception of 3-D form. *J. Exp. Psychol.* 45, 360-368.

Wallach, H., Weisz, Alexander and Adams, P. A. 1956 Circles and derived figures in rotation. *Am. J. Psychol.* 69 48-59.

White, B. W. and Mueser G. E. 1960 Accuracy in reconstructing the arrangement of elements generating kinetic depth displays. *J. Exp. Psychol.* 60(1), 1-11.

Zegers, R. T. 1964 The reversal illusion of the Ames trapezoid. *Trans. N.Y. Acad. Sci.*, 26, 377-400.

## APPENDIX

**The structure from motion theorem:**

Given three distinct orthographic projections of four non-coplanar points in a rigid configuration, the structure and motion compatible with the three views are uniquely determined up to a reflection about the image plane.

Comment: It is assumed that a *correspondence* between the projections has already been established. Namely it is known which points in the three views are the projection of the same source point in space.

Nomenclature: Let O, A, B, and C be the four points. The motion of the object is composed of translation and rotation. In orthographic projection the recovery of the translation in depth is impossible, and the recovery of the remaining translation component is trivial, since it is congruent in space and in the image plane. It is also assumed that corresponding points (i.e. the three projections of the same 3-D point) have been identified. The problem is therefore equivalent to the following formulation. The orthographic projections of the four points on three distinct planes PI, P2, P3, are given, and the 3-D configuration of the points is to be reconstructed. We choose a fixed coordinate system with its origin at O. Let $\underline{a}$ $\underline{b}$ and $\underline{c}$ be the vectors from O to A, B, and C, respectively. Let each view have a 2-D coordinate system $(p_i, q_i)$, with the image of O at its origin. That is, $\underline{p}_i$ and $\underline{q}_i$ are orthogonal unit vectors on Pi. Let the coordinates of (A, B, C) on Pi (the image coordinates) be $(x_{ai}, y_{ai}, x_{bi}, y_{bi}, x_{ci}, y_{ci})$ for i = 1,2,3. Finally, let $\underline{u}_{ij}$ be a unit vector along the intersection line of Pi and Pj.

The image coordinates are given by:

(1) $x_{ai} = \underline{a}\,\underline{p}_i \quad y_{ai} = \underline{a}\,\underline{q}_i$

$x_{bi} = \underline{b}\,\underline{p}_i \quad y_{bi} = \underline{b}\,\underline{q}_i$

$x_{ci} = \underline{c}\,\underline{p}_i \quad y_{ci} = \underline{c}\,\underline{q}_i$

The unit vector $\underline{u}_{ij}$ is on Pi which is spanned by $(\underline{p}_i, \underline{q}_i)$ hence:

(2) $\underline{u}_{ij} = \alpha_{ij}\underline{p}_i + \beta_{ij}\underline{q}_i, \quad \alpha_{ij}^2 + \beta_{ij}^2 = 1.$

The unit vector $\underline{u}_{ij}$ is also on Pj which is spanned by $(\underline{p}_j, \underline{q}_j)$ hence:

(3) $\underline{u}_{ij} = \gamma_{ij}\underline{p}_j + \delta_{ij}\underline{q}_j, \quad \gamma_{ij}^2 + \delta_{ij}^2 = 1.$

From (2) and (3) we get the vector equation:

(4) $\alpha_{ij}\underline{p}_i + \beta_{ij}\underline{q}_i = \gamma_{ij}\underline{p}_j + \delta_{ij}\underline{q}_j$

Taking the scalar product of (4) with $\underline{a}$, $\underline{b}$, and $\underline{c}$ respectively, we get:

(5) $\alpha_{ij} x_{ai} + \beta_{ij} y_{ai} = \gamma_{ij} x_{aj} + \delta_{ij} y_{aj}$

$\alpha_{ij} x_{bi} + \beta_{ij} y_{bi} = \gamma_{ij} x_{bj} + \delta_{ij} y_{bj}$

$\alpha_{ij} x_{ci} + \beta_{ij} y_{ci} = \gamma_{ij} x_{cj} + \delta_{ij} y_{cj}$

These three equations in $(\alpha_{ij}, \beta_{ij}, \gamma_{ij}, \delta_{ij})$ are linearly independent. For assume that there exist three scalars $\nu$, $\lambda$ and $\mu$, such that:

(6) $\nu\underline{p}_i\,\underline{a} + \lambda\underline{p}_i\,\underline{b} + \mu\underline{p}_i\,\underline{c} = 0$

$\nu\underline{q}_i\,\underline{a} + \lambda\underline{q}_i\,\underline{b} + \mu\underline{q}_i\,\underline{c} = 0$

$\nu\underline{p}_j\,\underline{a} + \lambda\underline{p}_j\,\underline{b} + \mu\underline{p}_j\,\underline{c} = 0$

$\nu\underline{q}_j\,\underline{a} + \lambda\underline{q}_j\,\underline{b} + \mu\underline{q}_j\,\underline{c} = 0$

In which case the vector $\underline{\theta} = (\nu\underline{a} + \lambda\underline{b} + \mu\underline{c})$ would be orthogonal to all of $\underline{p}_i$, $\underline{p}_j$, $\underline{q}_i$, and $\underline{q}_j$. Since Pi and Pj are distinct, this implies that $\underline{\theta} = \underline{0}$. But since (O, A, B, C) are non

coplanar, $\underline{\theta} = \eta\underline{a} + \lambda\underline{b} + \mu\underline{c} = \underline{0}$ implies that $\eta = \lambda = \mu = 0$, hence the equations are linearly independent.

Combined with the requirement $\alpha_{ij}^2 + \beta_{ij}^2 = \gamma_{ij}^2 + \delta_{ij}^2 = 1$, equations (5) have two solutions of the same absolute values but opposite signs. Choosing one of these solutions, we have found (up to a sign) the vectors $\underline{u}_{ij}$ in terms of $(\underline{p}_i, \underline{q}_i)$ and $(\underline{p}_j, \underline{q}_j)$.

We can next determine the distances:

(7) $d_1 = \| \underline{u}_{12} - \underline{u}_{13} \|$

$d_2 = \| \underline{u}_{12} - \underline{u}_{23} \|$

$d_3 = \| \underline{u}_{13} - \underline{u}_{23} \|$

We now examine the triangle whose sides are $d_1$, $d_2$, $d_3$. If there is a solution to the reconstruction problem then there exists at least one such triangle. It might, however, be degenerate, i.e. at least one of the the distances equals zero. In the non-degenerate case the triangle is unique, and all its vertices are known to lie at a unit distance from the origin. The three vertices and the origin thus define two possible tetrahedra, one being the reflection of the other. For each tetrahedron, the projections of A, B, and C on the three planes are known, and they determine a unique 3-D configuration.

The degenerate case: If one of the distances $d_i = 0$ ($i = 1,2,3$), then all are, and the tetrahedron degenerates to a single line. Otherwise, two of the planes $P_i$ ($i = 1,2,3$) would coincide, contrary to the assumption. To prove the degenerate case we shall first establish two lemmas.

Let $p_1 = (x_1\ y_1)$, $p_2 = (x_2\ y_2)$, $q_1 = (x'_1\ y'_1)$, $q_2 = (x'_2\ y'_2)$ be four points in a plane $(x\ y)$ such that $(q_1\ q_2)$ is not the reflection of $(p_1\ p_2)$ about the y-axis, and suppose that the lines $p_1$ -

$q_1$ and $p_2 - q_2$ (which we shall call "trajectories") are parallel to the x-axis. The pair $q_1 - q_2$ is now rotated by $\alpha$ about the origin O. Let $q'_1 = (u_1\ v_1)$ and $q'_2 = (u_2\ v_2)$ be the rotated pair. O, $p_1$, and $p_2$, are assumed to be non-collinear.

Lemma 1: There exists exactly one angle $\alpha > 0$ such that the lines $p_1 - q'_1$ and $p_2 - q'_2$ are parallel.

Proof: For $i = 1,2$:

(8) $u_i = x'_i \cos\alpha - z'_i \sin\alpha$,

$v_i = x'_i \sin\alpha + z'_i \cos\alpha$

If, following the rotation, the lines are parallel, then their slopes coincide, namely:

(9)

$$\frac{y_1 - v_1}{x_1 - u_1} = \frac{y_2 - v_2}{x_2 - u_2}$$

(There is also the possibility that the two lines are parallel to the y-axis in which case the denominators in both the above expressions vanish. We shall see, however, that in this case the solution is still unique.)

Substituting for v and w using (8) yields: ($y'_1 = y_1$ and $y'_2 = y_2$)

(10)

$$\frac{y_1 - x'_1\sin\alpha - y_1\cos\alpha}{x_1 - x'_1\cos\alpha + y_1\sin\alpha} = \frac{y_2 - x'_2\sin\alpha - y_2\cos\alpha}{x_2 - x'_2\cos\alpha + y_2\sin\alpha}$$

Which reduces to the form:

(11)

$$a \sin\alpha + b \cos\alpha = b \qquad \text{where:}$$

$$a = x_1x'_2 - x'_1x_2$$

$$b = x_1y_2 + x'_1y_2 - y_1x_2 - y_1x'_2$$

For given a and b this equation has exactly one solution for $\alpha$, given by:

(12)

$$\sin\alpha = 2ab / (a^2 + b^2)$$

$$\cos\alpha = (b^2 - a^2) / (a^2 + b^2)$$

The only case where there is no unique solution to (12) is when both a and b are equal to 0. In this latter case (11) provides two equations in $x'_1$ and $x'_2$. If the equations are independent, then their solution is: $x'_1 = -x_1$, $x'_2 = -x_2$ namely, $(q_1\ q_2)$ is the reflection of $(p_1\ p_2)$ about the y-axis, contrary to the assumption. If they are independent then $x_1/x_2 = y_1/y_2 = x'_1/x'_2$ which violates the non-colinearity assumption. If there exists a rotation $\beta$ which makes the denominators in (9) equal to zero, this $\beta$ is still a solution -- and the only solution -- to equation (11). ∫

Lemma 2: If the projections of two objects O and O' on the frontal plane coincide, and if the coincidence is maintained after both objects rotate by the same amount y ($y < 180$ degrees) about the vertical axis, then O and O' are congruent.

The proof is straightforward and will be omitted.

We wish to establish the uniqueness of the interpretation for (O, A, B, C) rotating about a fixed axis. Let the rotation axis be the z-axis of a coordinate system whose origin is at O, and y-z be the image plane. Let $\Omega$ be the object (O, A, B, C). If the interpretation is not unique then an object W' = (O, A', B', C') exists whose rotations are different than those of $\Omega$, and the three projections of W and W' coincide. (By lemma 2, if the rotations are the same the objects are congruent).

Between the first and the second views, $\Omega$ rotated by some angle $\alpha_1 \neq 0$, and W' by $\beta_1$. Between the second and third views, $\Omega$ rotated by $\alpha_2 \neq 0$, $\Omega'$ by $\beta_2$. Throughout the rotations the projections of $\Omega$ and $\Omega'$ on the image plane y-z coincide. Let $p_1 = (x_1\ y_1)$ be the projection of A on the x-y plane, $p_2 = (x_2\ y_2)$ the projection of B, $q_1 = (x'_1\ y'_1)$ the projection of A', and $q_2 = (x'_2\ y'_2)$ the projection of B'. Without loss of generality $(p_1, p_2)$ and $(q_1, q_2)$ satisfy the requirements of Lemma 1 (since two such pairs must exist, if not the projections of A, B, A', B', then the projections of A, C, A', C').

claim: $\alpha_2 = \beta_2$

Proof of the claim: Between the first and second view $\Omega$ (and so $p_1\ p_2$) rotated by $\alpha_1$ and $\Omega'$ (and so $q_1\ q_2$) rotated by $\beta_1$, and the resulting trajectories $p_1 - q_1$ and $p_2 - q_2$ remain parallel to the x-axis. If $\Omega$ does not rotate, and $\Omega'$ rotates by $\beta_1 - \alpha_1$, then the trajectories will still be parallel to each other (though not to the x-axis). According to lemma 1 there is a unique angle for which this will happen. Call this angle y, then $\beta_1 - \alpha_1 = y$. Between the first and third view $\Omega$ rotated by $\alpha_1 + \alpha_2$ and $\Omega'$ by $\beta_1 + \beta_2$ resulting in parallel trajectories. Therefore $(\beta_1 + \beta_2) - (\alpha_1 + \alpha_2) = y$. We get:

(13) $\beta_1 + \beta_2 - \alpha_1 - \alpha_2 = \gamma$

But since: $\beta_1 - \alpha_1 = \gamma$, $\beta_2 = \alpha_2$.

Between the second and third view the two objects retain their coincidence of projection through a common rotation. According to lemma 2 they are congruent. ∎

The above proof offers a way of actually computing 3-D structure from three orthographic projections. The computation has to be expressed in terms of the measurable parameters, which are the 2-D coordinates of the four points in the three views expressed in terms of $(p_i, q_i)$ for $i = 1,2,3$. Equations (5) use these parameters to determine $\underline{u}_{ij}$, the unit vectors generating the tetrahedron. If the tetrahedron is non-degenerate, two views are sufficient to determine the 3-D configuration. The 3-D position of a point can be found by the intersection of the perpendiculars to its projections on two planes. The recovery of the structure in the degenerate case is not given by the proof but can be determined by straightforward trigonometric considerations [Ullman 1977, Appendix 2].